CLAIMS

What Is Claimed Is:

1. A processing apparatus comprising:

structured/hierarchical content, said apparatus makes a determination whether or not said structured/hierarchical content delivered through a network includes a content portion matched with a predetermined matching pattern, and performs predetermined processing for the structured/hierarchical content if a result of the determination is positive, the processing apparatus further comprising:

target subtree setting means for setting a target subtree relating to a range including a target content portion as an extracted portion of the matching pattern in the structured/hierarchical content (hereinafter, referred to as a "target content") from which the matching pattern is to be extracted;

occurrence mode detecting means for detecting an occurrence mode of each node of the target subtree by selecting a plurality of past structured/hierarchical

contents with respect to the target content and collating the target subtree relating to the target content with a tree relating to each of the past structured/hierarchical contents;

statistical information generating means for generating statistical information concerning an occurrence frequency of the occurrence mode of each node in the target subtree based on the plurality of past structured/hierarchical contents;

classifying means for performing classification of each node of the target subtree based on the statistical information and a result of detecting the occurrence mode; and

matching pattern generating means for generating the matching pattern for the target content portion based on the classification.

2. The processing apparatus according to claim 1, wherein the predetermined processing is to associate related information with the content portion of the structured/hierarchical content.

3. The processing apparatus according to claim 2, wherein the related information includes an annotation.

4. The processing apparatus for a structured/hierarchical content according to claim 1, wherein the predetermined processing is processing for copying the content portion of the structured/hierarchical content for a purpose of utilizing the content portion of the structured/hierarchical content for another structured/hierarchical content.

5. The processing apparatus for a structured/hierarchical content according to claim 1, wherein the structured/hierarchical content is a Web content.

6. The processing apparatus for a structured/hierarchical content according to claim 1, wherein the classifying means classifies nodes of the target subtree into stationary nodes, updated nodes and additional nodes.

7. The processing apparatus according to claim 6, wherein the occurrence mode detecting means includes, as the occurrence mode to be detected, (N1) an occurrence mode

where detected nodes occur in both of the target content portion and structured/hierarchical contents collated therewith and contents thereof are mutually identical, and (N2) an occurrence mode where the detected nodes occur in both of the target content portion and the structured/hierarchical contents collated therewith and the contents thereof are mutually different, and

the classifying means classifies, into the stationary nodes, nodes of which occurrence frequency of the occurrence mode (N1) is determined to be equal to/more than a first threshold value by the statistical information, classifies, into the updated nodes, nodes of which occurrence frequency of the occurrence mode (N2) is determined to be equal to/more than a second threshold value by the statistical information, and classifies, into the additional nodes, nodes other than the stationary nodes and the updated nodes.

8. The processing apparatus according to claim 6, wherein the matching pattern generating means includes:

repeated portion detecting means for detecting a repeated portion in the target subtree based on the classification into the stationary nodes, the updated nodes

and the additional nodes; and

repeated information-added matching pattern generating means for generating the matching pattern including presence information of the repeated portion.

9. The processing apparatus according to claim 8, wherein the classifying means includes:

formed-for-spacer image detecting means for detecting whether or not a node relating to an image is a node relating to a formed-for-spacer image for ensuring a blank region;

bullet image detecting means for detecting whether or not the node relating to the image is a node relating to a plurality of bullet images used repeatedly in a same size;

first classifying means for classifying the node relating to the formed-for-spacer image into the additional nodes; and

second classifying means for allocating a plurality of the nodes relating to the bullet image into a same classification among classifications of the stationary nodes, updated nodes and additional nodes even if display contents of the plurality of nodes are mutually different.

10. The processing apparatus according to claim 1, further comprising:

collating means for collating the target subtree relating to the target content with the trees relating to a plurality of structured/hierarchical contents adjacent to the target content by selecting the adjacent structured/hierarchical contents in place of the past structured/hierarchical contents with respect to the target content when the past structured/hierarchical contents are not present.

11. A processing apparatus comprising:

structured/hierarchical content, said apparatus makes a determination whether or not said structured/hierarchical content delivered through a network includes a content portion matched with a predetermined matching pattern, and performs predetermined processing for the structured/hierarchical content if a result of the determination is positive, the processing apparatus comprising:

target subtree setting means for setting a target

subtree relating to a range including a target content portion as an extracted portion of the matching pattern in the structured/hierarchical content (hereinafter, referred to as a "target content") from which the matching pattern is to be extracted;

occurrence mode detecting means for detecting an occurrence mode of each node of the target subtree by selecting a plurality of structured/hierarchical contents adjacent to the target content and collating the target subtree relating to the target content with a tree relating to each of the adjacent structured/hierarchical contents;

statistical information generating means for generating statistical information concerning an occurrence frequency of the occurrence mode of each node in the target subtree based on the plurality of adjacent structured/hierarchical contents;

classifying means for performing classification of each node of the target subtree based on the statistical information and a result of detecting the occurrence mode; and

matching pattern generating means for generating the matching pattern for the target content portion based on the

classification.

12. A processing method comprising:

making a determination whether or not structured/hierarchical content delivered through a network includes a content portion matched with a predetermined matching pattern, and performs predetermined processing for the structured/hierarchical content if a result of the determination is positive, the processing method further comprising:

a target subtree setting step of setting a target subtree relating to a range including a target content portion as an extracted portion of the matching pattern in the structured/hierarchical content (hereinafter, referred to as a "target content") from which the matching pattern is to be extracted;

an occurrence mode detecting step of detecting an occurrence mode of each node of the target subtree by selecting a plurality of past structured/hierarchical contents with respect to the target content and collating the target subtree relating to the target content with a tree relating to each of the past structured/hierarchical

contents;

a statistical information generating step of generating statistical information concerning an occurrence frequency of the occurrence mode of each node in the target subtree based on the plurality of past structured/hierarchical contents;

a classifying step of performing classification of each node of the target subtree based on the statistical information and a result of detecting the occurrence mode; and

a matching pattern generating step of generating the matching pattern for the target content portion based on the classification.

13. The processing method according to claim 12, wherein the predetermined processing is to associate related information with the content portion of the structured/hierarchical content.

14. The processing method according to claim 13, wherein the related information includes an annotation.

15. The processing method according to claim 12, wherein the predetermined processing is processing for copying the content portion of the structured/hierarchical content for a purpose of utilizing the content portion of the structured/hierarchical content for another structured/hierarchical content.

16. The processing method according to claim 12, wherein the structured/hierarchical content is a Web content.

17. The processing method according to claim 12, wherein in the classifying step, nodes of the target subtree are classified into stationary nodes, updated nodes and additional nodes.

18. The processing method according to claim 17, wherein the occurrence mode detecting step includes, as the occurrence mode to be detected, (N1) an occurrence mode where detected nodes occur in both of the target content portion and structured/hierarchical contents collated therewith and contents thereof are mutually identical, and (N2) an occurrence mode where the detected nodes occur in both of

the target content portion and the structured/hierarchical contents collated therewith and the contents thereof are mutually different, and

in the classifying step, are classified into the stationary nodes, nodes of which occurrence frequency of the occurrence mode (N1) is determined to be equal to/more than a first threshold value by the statistical information, are classified into the updated nodes, nodes of which occurrence frequency of the occurrence mode (N2) is determined to be equal to/more than a second threshold value by the statistical information, and are classified into the additional nodes, nodes other than the stationary nodes and the updated nodes.

19. The processing method according to claim 17, wherein the matching pattern generating step includes:

a repeated portion detecting step of detecting a repeated portion in the target subtree based on the classification into the stationary nodes, the updated nodes and the additional nodes; and

a repeated information-added matching pattern generating step of generating the matching pattern including

presence information of the repeated portion.

20. The processing method according to claim 19, wherein the classifying step includes:

a formed-for-spacer image detecting step of detecting whether or not a node relating to an image is a node relating to a formed-for-spacer image for ensuring a blank region;

a bullet image detecting step of detecting whether or not the node relating to the image is a node relating to a plurality of bullet images used repeatedly in a same size;

a first classifying step of classifying the node relating to the formed-for-spacer image into the additional nodes; and

a second classifying step of allocating a plurality of the nodes relating to the bullet image into a same classification among classifications of the stationary nodes, updated nodes and additional nodes even if display contents of the plurality of nodes are mutually different.

21. The processing method according to claim 12, further comprising:

a collating step of collating the target subtree relating to the target content with the trees relating to a plurality of structured/hierarchical contents adjacent to the target content by selecting the adjacent structured/hierarchical contents in place of the past structured/hierarchical contents with respect to the target content when the past structured/hierarchical contents are not present.

22. A processing method comprising:

making a determination whether or not structured/hierarchical content delivered through a network includes a content portion matched with a predetermined matching pattern, and performs predetermined processing for the structured/hierarchical content if a result of the determination is positive, the processing method comprising:

a target subtree setting step of setting a target subtree relating to a range including a target content portion as an extracted portion of the matching pattern in the structured/hierarchical content (hereinafter, referred to as a "target content") from which the matching pattern is to be extracted;

an occurrence mode detecting step of detecting an occurrence mode of each node of the target subtree by selecting a plurality of structured/hierarchical contents adjacent to the target content and collating the target subtree relating to the target content with a tree relating to each of the adjacent structured/hierarchical contents;

a statistical information generating step of generating statistical information concerning an occurrence frequency of the occurrence mode of each node in the target subtree based on the plurality of adjacent structured/hierarchical contents;

a classifying step of performing classification of each node of the target subtree based on the statistical information and a result of detecting the occurrence mode; and

a matching pattern generating step of generating the matching pattern for the target content portion based on the classification.

23. A computer program product comprising a computer usable medium having computer readable program code means embodied therein for causing functions of the apparatus, the computer

readable program code means in said computer program product comprising computer readable program code means for causing a computer to effect the functions of claim 1.

24. A computer program product comprising a computer usable medium having computer readable program code means embodied therein for causing functions of the apparatus, the computer readable program code means in said computer program product comprising computer readable program code means for causing a computer to effect the functions of claim 11.

25. An article of manufacture comprising a computer usable medium having computer readable program code means embodied therein for causing processing, the computer readable program code means in said article of manufacture comprising computer readable program code means for causing a computer to effect the steps of claim 12.

26. A program storage device readable by machine, tangibly embodying a program of instructions executable by the machine to perform method steps for processing, said method steps comprising the steps of claim 12.

27. An article of manufacture comprising a computer usable medium having computer readable program code means embodied therein for causing processing, the computer readable program code means in said article of manufacture comprising computer readable program code means for causing a computer to effect the steps of claim 22.

28. A program storage device readable by machine, tangibly embodying a program of instructions executable by the machine to perform method steps for processing, said method steps comprising the steps of claim 22.